

Chapter 2

Credible Methodology

Abstract The goal of research is to generate evidence, from which people can learn and can base their decisions. There are persistent debates among evaluators about the credibility of evidence and the methodologies and methods used to generate it. Justifying the appropriateness of methods and explaining assumptions about their validity is a critical part of methodological credibility. Validity comes from the accuracy of methods in depicting reality and the conclusions and decisions based on this depiction. In this chapter, tenets of a credible evaluative argument are discussed.

Introduction

Research credibility can be defined as the ability of a research process to generate findings that elicit belief and trust (O’Leary, 2007). Two research paradigms, the positivist (scientific) and post-positivist (qualitative), illustrate how researchers perceive their role and the nature of truth. The goal is the same for both paradigms—to generate new knowledge from which people can learn and on which they can base decisions.

In the positivist approach, social science research follows the rules of the natural sciences, with indicators premised on a knowable world that can be measured objectively and quantifiably (O’Leary, 2007). The world is rational and patterned, with systems that can be understood and generalized to populations at large, if studied under a careful research structure. To a post-positivist, the world is ambiguous, complex, constructed, and open to interpretation. Researchers who work under qualitative or post-positivist frameworks see themselves as subjective beings and therefore, attempt to make clear any initial biases they may have toward the work they are doing.

One of the greatest challenges for researchers, particularly those comfortable with the positivist paradigm, is to have their work seen as credible. Research must be transparent and reliable to be credible, and results will only be incorporated in a body of knowledge if they are considered accurate and worthwhile.

Credibility of Evidence, Methodologies, and Methods

There are persistent debates among evaluators about credible evidence. Two examples of such debates are pointed out here:

- a. the Claremont Symposium,
- b. the 2009 Australasian Evaluation Society conference expert panel on hierarchies of evidence

The Claremont Symposium Donaldson, Christie, and Mark's 2009 book "What Counts as Credible Evidence in Applied Research and Evaluation Practice?" synthesizes discussions in an "illuminating and action packed" (p. vii) 2006 debate on the subject among more than 200 participants in the Claremont symposium (www.cgu.sbos).

Australasian Evaluation Society Conference Expert Panel on Hierarchies of Evidence Hierarchies of evidence were debated during a 2009 Australasian Evaluation Conference in Canberra. The reader may be interested in the edited transcript of that discussion at the end of this chapter (Annex 2.1).

Both debates focused on how to judge the quality of evidence, but participants discussed the methodologies and methods used to generate evidence (e.g., randomized control evaluations; methodological sophistication, rigor, and use; application of mixed methods; and quantitative versus qualitative methods). The 2009 Australasian Evaluation Society conference expert panel on hierarchies of evidence concluded that

- Randomized control evaluations seek to measure a counterfactual, but understanding the "factual" to which the "counter" is sought is even more critical. This is related to the construct of the program or evaluand—the evaluators' understanding of its outcomes, functioning, etc.
- Randomized control evaluations need strong program descriptions.
- Methodological rigor should not only be limited to the appropriate use of methods, but also includes sufficient understanding of the problem or program to which the tools or methods are applied, the right timing and sequence of evaluation.
- It is more appropriate to talk about cycles of different hierarchies of evidence that are linked to the types of evaluation questions rather than a pyramidal hierarchy of evidence. That is, there may be a hierarchy for descriptive questions under certain conditions.

In reference to this debate about evidence, Donaldson et al. remarked "we fear that history has been repeating itself of late, with the earlier paradigm war appearing again in varied guises." (2009, p. ii). Essentially, methodology is the recurrently dominant element in discussions about evidence. Though closely related, evidence and methodology are quite different. Evidence is the information needed to form a conclusion or determine whether a proposition is true (Schwandt 2005). Methodology was defined in the previous chapter as the application of theory in the form of specific and coherent steps, procedures, and measures. The nature of the

evidence needed has a direct bearing on the research methodology used, and vice versa. Credibility applies to both evidence and the methods by which the evidence arises. At the minimal, methods need to ensure validity.

Validity is the extent to which appropriate conclusions, inferences, and actions are derived from measurement and research. It is a common notion in reference to measurement, but it is much broader than that. It has to do with whether or not and how the purposes of research and measurement are correctly derived (House 1977), whether findings reflect what is researched or measured (Lipsey 1988), and whether appropriate research methods and measures support the interpretation of data and decisions made (Hoyt et al. 2006; House 1977; Lipsey 1988; Messick 1989; Kirkhart 2010; Reichardt 2011). An evaluation design or set of measures may contribute to more or less valid conclusions, interpretations, and decisions under different circumstances (Reichardt 2011). The valid use of research findings may depend somewhat on the statistical connectedness of data, but is ultimately a function of much broader issues, such as stakeholder values (House 1977).

Lipsey's 1988 analogy between drawing and evaluation illustrates this point. An artist who attempts to translate an experience, such as beautiful scenery, into a drawing must not distort the essential features and patterns of the scenery. If the drawing misrepresents the original experience, our resultant understanding of the artist's experience lacks validity. The artist's process, style, and tools may affect our translation of experience and are important for determining its perceived validity. More importantly, still is, whether the drawing helps us understand the original experience, the conclusions we make about the original experience based on the drawing, and the decisions we make based on such conclusions.

Similarly, evaluators who work with complex and dynamic programs, stakeholders, and contexts have to design measures (methods, variables, constructs, and comparisons) that represent the important dynamics, effects, and elements of the programs (including participants) *as accurately as possible* (author's emphasis). Validity depends on the translation of the measures, the truthfulness in the depiction of the program situation (affected by the accuracy of measures), and the conclusions and decisions based on this translation.

Arguments that justify the appropriateness of methods are a critical part of methodological credibility, and explaining assumptions about the validity of selected methods are invariably a critical part of the methodological credibility argument. Methodology is understood to be superior to and broader than methods. Kaplan (1964, p. 18) argues that methodology "is the study—the description, the explanation and justification—of methods" ... is "to describe and analyze these methods, throwing light on their limitations and resources, clarifying their presuppositions and consequences, relating their potentialities to the twilight zone at the frontiers of knowledge" (p. 23).

This definition relates methodology to arguments about applied methods. If methodology is about arguments, methodological credibility is about the quality of those arguments. Ernest House's work on validity emphasizes the centrality of argument in methodology (1977, 2014). House broadened the meaning of validity beyond addressing biases in research design and credibility (Montrosse-Moorhead, Griffith and Pokorny, 2014, p. 97) to include truth, beauty, and justice. "Truth is the attainment of arguments soundly made, beauty is the attainment of coherence well-

wrought, and justice is the attainment of politics fairly done (Montrosse-Moorhead et al. 2014, p. 97).” Within these validity criteria, credibility is aligned with truth. Truth is not about “the designs or correlations that determined validity, but rather the validity of the arguments that supported the use of the designs and correlations” (House, 2014, p. 9).

At the time House wrote this (around 1977, based on references used), evaluation quality was more or less defined in terms of validity. Evaluations were expected to address biases through various experimental designs to provide evidence that was certain. This required “isolating data from its local context” (House, 1977, p. 5). House argued that the context in which decisions, even personal decisions, are made and subsequent actions are taken is complex. There is an indirect relation between information and actions, that what precedes the actions is not the information itself but arguments about that information, that social phenomena themselves are complex and difficult to know with certainty, and that “in the face of uncertain knowledge, the task of entangled decision-making becomes less one of absolutely convincing ourselves with proofs than one of persuading ourselves with multiple reasons.” “The criterion becomes not what is necessary but what is plausible” (House, 1977, p. 2). House saw argumentation as the main tool of persuasion, to win “a particular audience to a point of view or course of action by appeal to the audience’s reason and understanding” (House, 1977, p. 5). As opposed to demonstration, which is based on avoidance of bias and pursuit of certain knowledge, argumentation focuses on what is credible, plausible, and probable (House, 1977). In 2014, House added that all validity types, including beauty and justice, inescapably rely on supporting arguments. The “beauty” element of validity refers to aesthetic qualities such as coherence and appropriate forms of communication that support the credibility of an evaluation argument (Bledsoe, 2014).

In explaining credibility, House refers to Joseph Gusfield’s 1976 synthesis of drunk driver studies, which found that contrary to what most people thought, drunk driving was a lot riskier than social drinking (Cited in Montrosse-Moorhead et al. 2014, p. 97).

The studies Gusfield reviewed changed the image of the drunk driver to one of a falling down drunk. So, now you have this image, which holds the whole thing together in a vivid way. It is very believable, highly credible if you bought into the image, and the image acted coherently to hold that together. You could conceivably look at the studies Gusfield reviewed and see if they really supported the idea of the drinking driver as being a falling down drunk. If that were not the case, using the image in that way would be inappropriate. And that would be an invalid conclusion about those studies.

House explains that credibility is about the clarity, coherence, and perhaps cogency with which evaluation findings or evidence are portrayed and framed as well as the accuracy of those findings. He maintained that judgment about “image,” how credibly evaluation results are portrayed, can be influenced by both implicit and explicit cognitive biases.

According to Davidson (2014), well-crafted and coherent evaluative arguments are not merely aesthetic. They unlock the door to stakeholder understanding and

evaluative reasoning and produce robust and defensible answers that are both credible and valid. Davidson describes a credible evaluative argument as unearthing assumptions, e.g., about values, justice principles, and the *reasoning used to arrive at robust conclusions* (author's emphasis). House, too, argues that assumptions are a critical element of evaluation: "actually many assumptions lie concealed behind the numbers, as indeed behind every evaluation" (House, 1977, p. 11). Therefore, a substantial part of evaluative argument is elaborating and challenging such assumptions: "one can almost challenge the validity of the tests, the appropriateness of the statistical procedures, and the control of experimental design. The challenge does not invalidate the evaluation, but once the premises are challenged, the nature of the evaluation as argumentation becomes apparent. The evaluator may defend his study either successfully or unsuccessfully" (House, 1977, p. 11).

Researchers such as Toulmin (1964) and Fournier (1995 a, b) have discussed approaches or argumentation that help unearth assumptions. These are discussed in the following section.

Tenets of a Credible Evaluative Argument

Argument is defined by Blair (1995, p. 72) as tools of persuasion, as well as the reasons one provides in support of a claim. Evaluative argument is essential to validity. A broad and interpretive vision of validity—one that transcends methodological and procedural accuracy—recognizes that what evaluators bring to an evaluation process is much more than the understanding of how to evaluate. According to Greene (2011), an evaluator is an "interested social inquirer"; "one who inevitably brings to the process of social inquiry his or her own sociocultural history, beliefs about the social world and about what constitutes warranted knowledge of it, theoretical preferences, and moral and political values. There is no location outside one's own self—no "view from nowhere"—that permits an objective and disinterested view of social phenomena, their interrelationships, and their meanings." (Greene, 2011, p. 82). Thus, "the reasoning behind evaluative claims is vital because at any one time, the view point, that is the most objective, is the one that currently is the most warranted or rational...." (Phillips, 1992, p. 119; cited in Fournier 1995b, p. 16).

An evaluative argument seeks not only to generate inferences from procedural evidence, but also to generate conclusions that are "meaningful, plausible, and of some consequence in the contexts at hand" (Greene, 2011, p. 82). Evaluative arguments are formulated by evaluative reasoning—the study of systematic means for arriving at evaluative conclusions and the principles that support inferences drawn by evaluators (Fournier, 1995a). The tenets of a credible evaluative argument include among others: addressing contextual issues, addressing relationships and values, getting to the key messages, and articulating and communicating the key messages.

Addressing Contextual Issues

Conclusions need to clarify the interaction between contextual factors and the program elements being evaluated (Greene, 2011; Rog, 1995). The contexts, as well as evaluands, are complex with a variety of influences. The evaluator engages with these different influences in ways that necessitate the evaluator to rely on more than standardized protocols. These contextual influences on the program, as well as process, need to be a part of a credible argument. For example, an evaluation of a community development program in Southern Ethiopia revealed that the percent of participating households owning larger pieces of land (above 1 ha) had declined from 22.1 % in 1996 to 11.5 % in 2012. At the same time, the percentage of households that owned private land had increased dramatically from 14 to 98.1 % in the same period of time. Contextual analysis indicated that these changes had more to do with the government land redistribution policies than the program's economic empowerment activities (World Vision, 2012).

Addressing Relationship Issues

According to Greene (2011, p. 83), "evaluation is fundamentally a social and relational craft." The inquiry process involves multiple relationships and numerous communications with multiple kinds of stakeholders. These interactions shape not only the process, but also the product of inquiry. The important relational appeals and how they potentially bear on findings need to be articulated within an evaluative conclusion. For example, a researcher investigating factors that influenced fathers' involvement in childcare in Central Uganda organized separate focus group discussions with mothers and fathers. However, fathers insisted on participating in the mothers' focus group discussions, perhaps as way to censor what they would say. The researcher agreed to these terms. The focus group discussions turned out to be much more dynamic and revealing about the situation of gender roles than those for fathers or mothers alone. In presenting conclusions that emerged from focus groups discussions, this unanticipated relational dynamic needed to be clarified (Nkwake, 2013).

Addressing Stakeholder Values

The criteria used to judge merit or worth are always premised on stakeholder values. "An inference can gain strength through an argument that references the value dimensions of the program and policy being evaluated, that invokes the value strands of evaluative conversation held along the way, and that demonstrates respect for and

acceptance of multiple value claims and stances” (Greene, 2011, p. 86). Stakeholder values influence priorities for which evaluation questions answer and the standards in defining merit or worth of interventions. Differences in values may thus be a source of contention in interpreting evaluation findings. Since these values may be tacit, evaluative arguments do well to unearth and explicate such assumptions.

Getting to the Key Messages

Rog (1995, p. 96) uses the phrase “getting to the bottom line (s)” to refer to this product of evaluative reasoning—the key messages. This involves triangulating and examining the data for the most coherent answers to evaluation questions. This reflective activity is a part of what is called evaluative reasoning. Fournier (1995a) defines reasoning as a study of the systematic means for arriving at evaluative conclusions and the principles that support inferences drawn by evaluators. Evaluative reasoning is primarily concerned with “what and how inferences are made and supported (the inferential processes) and with the quality, weighing, and marshaling of evidence in building a case (the evidentiary processes)” (Fournier, 1995a, p. 1).

Articulating and Communicating the Key Message

This involves framing conclusions with sufficient clarity, providing explanations for controversial information. There is debate on the extent to which an evaluator should accompany his or her conclusions with recommendations. Some evaluators (e.g., Scriven, 1995; Rog, 1995) advise that evaluators should make recommendations with considerable caution. And Patton (1994) suggests that it is better for an evaluator to present findings and conclusions, and use these to facilitate discussions on appropriate recommendations by stakeholders. Whether recommendations are made from claims or not, arriving at bottom lines, as Fournier calls them, involves a significant amount of reasoning and application of evaluation logic.

Evaluation Logic

Scriven (1995) defines evaluation logic as the specific principles of reasoning that underlie the inference processes in all the fields of evaluation. There are some generally applied logics, such as inductive logic, deductive logic, and statistical inference; and yet, while these are commonly used in evaluation, “there is nothing evaluation-specific about them” (Scriven, 1995, p. 49).

General Logic

General logic refers to basic rules for constructing and testing claims. The kinds of reasoning necessary to justify a conclusion. According to Fournier (1995b), the general logic in evaluation is comprised of the following elements:

- Establishing a criteria for merit—the dimensions on which the worth or merit of an evaluand should be judged.
- Constructing standards—definitions of how well an evaluand should perform.
- Measuring the performance and comparing with the standards—establishing how well an evaluand has performed.
- Synthesizing and integrating data into a judgment of merit or worth—determining the merit and worth of a program.

This logic applies to all fields in evaluation evaluating an intervention in education, healthcare, governance, agriculture, etc. would apply this logic in one way or another (Fournier, 1995b).

Working Logic

Developing and applying criteria and standards for measuring worth (general logic) are applied in different ways from one evaluand to another. This tactical activity applies working logic (Fournier, 1995b). A nuanced logic emerges as an evaluator, tries to understand the evaluand (including its goals, issues being addressed, etc.) and its contexts as evaluation questions are conceptualized; appropriate evaluation designs are determined; measures are developed; data are collected and analyzed; and evaluative conclusions or claims are made. In complex programs, differences in working logic emerge across program components. For example, a program aimed at increasing the resilience of communities to shocks and stresses (component 1) also aims at increasing the generation and use of resilience data (component 2). Evaluating the first component might involve measuring in a range of the extent to which communities' resilience is being improved, yet the quality and utilization of these very measures may itself be an important criterion in evaluating the second component.

Another way form of work logic, according to Fournier (1995b) is the “argument structure that makes concrete the reasoning pattern by which conclusions are justified” (p. 23). According to Fournier, the structure of an evaluative argument involves six logical features that were identified by:

- a. *Claims*: the conclusion of what is taken to be acceptable and legitimate
- b. *Evidence*: the facts forming the basis for the claim
- c. *Warrants*: what legitimatizes the inferences drawn from evidence by appeal to some authority?
- d. *Backings*: what supports the warrant by appeal to some more general authority—why a warrant should be accepted as backing for the inference?

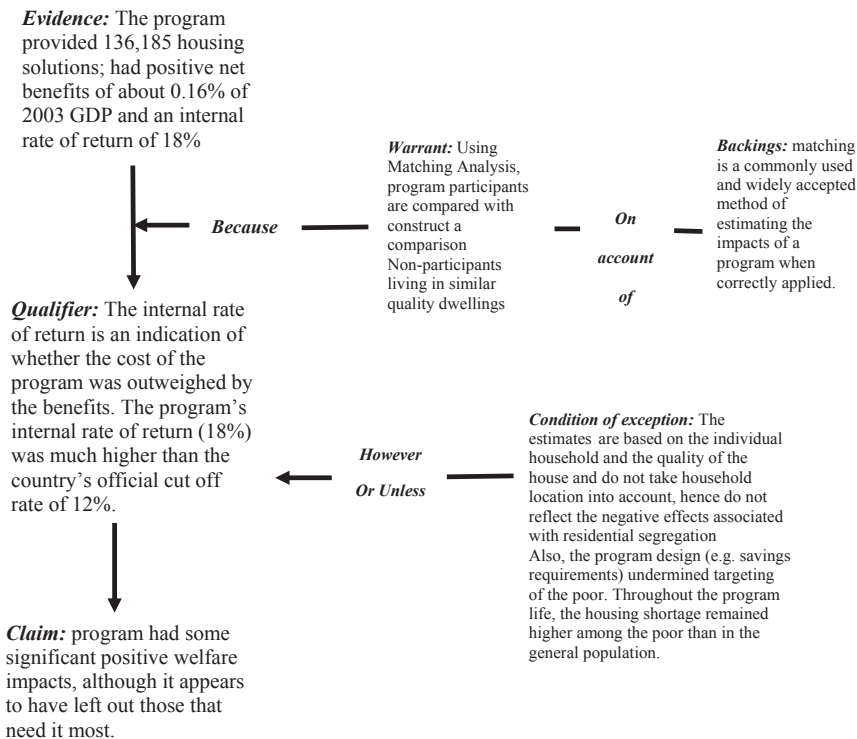


Fig. 2.1 An illustration of Toulmin's structure of an evaluative argument

- e. *Conditions of exception*: limitations or circumstances in which the warrant may not hold
- f. *Qualifiers*: identify the strengths of a claim

A summary of one claim was presented in the Chile Progressive Housing Program evaluation. This was a public housing program that facilitated the purchase of a new home (more information on this program and its evaluation findings can be found in (Marcano & Ruprah, 2008) (Fig. 2.1).

Within Toulmin's model, the qualifiers, warrants, backings, and conditions of exception are all nuances that are too often excluded from the presentation of evaluative claims. In the next chapters and sections of this book, these aspects are discussed as assumptions.

Assumptions

Assumptions compose one of the five epistemological notions which include: belief, justification, probability, truth, and assumptions (Sherman and Harman 2011). To assume is to take something to be true, in accordance with fact or reality and to

believe and accept that a thing exists (OECD 2002). To assume is to suppose, to take as given, to take for granted, or to conclude, deduce, or reason (Nkwake 2013). Therefore, it is difficult to exclude what we assume from what we consider known. For example, your knowledge that your car is outside where you parked it in front of your house is based on the assumption that the car has not been stolen, something you may not know, but *justifiably* take for granted (Sherman & Harman, 2011). This does not mean your car has not been stolen. Knowing something (X), which depends on justifiably taking another thing for granted (Y), does not imply with certainty that if X, then Y, although this is usually the case. It is easy to believe completely what you merely take for granted without evidence that it is true. You may also believe what you take for granted, but consider it a working hypothesis rather than the truth. Indeed, in the context of scientific inquiry (and therefore evaluation), a set of assumptions is a logical starting point for a causal explanation and produces a set of empirically testable hypotheses (Nkwake 2013; Kanazawa 1988).

Methodological Assumptions

Assumption is often accompanied by justification, which is the act of showing something to be right or reasonable (OECD 2002). To believe in something is to judge that it is right, probable, or likely (OECD 2002). The following issues are crucial for methodological assumptions:

- What is taken for granted about the validity of method decisions, which can differ when an evaluation is approached from a constructivist (subjective) vs. a positivist (objective) viewpoint
- Whether or not it is justifiable to take these things for granted (and how to ascertain this), and
- What implications does whatever is taken for granted (especially unjustifiably) have for evaluation quality and credibility?

All of these questions are influenced by evaluators' paradigmatic assumptions, what drives their perspectives on the reality in which they live. These assumptions can take many forms—ontological (thinking a certain way about the nature of reality) or idiosyncratic (shaping opinions based on personal experience). Evaluators structure their methodology according to the nature of their assumptions. Each kind of assumption should be explored to adjust for bias and avoid drawing unwarranted conclusions about methodology and data interpretation. These issues are discussed in as preconditions for validity at various stages of the evaluation process (Fig. 2.2). These preconditions need to be examined to ensure that validity is not compromised, i.e., assumptions arise when ever these preconditions are left unexamined or taken for granted. Credibility depends on the arguments that justify appropriateness of method choices, the extent to which preconditions for validity are met, and the ramifications for unmet preconditions.

Credibility, Validity, and Assumptions in Program
Evaluation Methodology

Nkwake, A.M.

2015, XXI, 166 p. 24 illus., 2 illus. in color., Hardcover

ISBN: 978-3-319-19020-4